

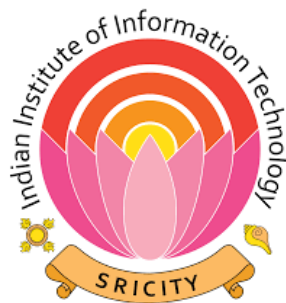
Multi Disease Prediction Framework

A BTP Report

by

Name	Roll No
Sivasanath Kumar Medavarapu	S20210020322
Sandeep Nidamanuri	S20210020300

Guide Name: Dr. Priyanka Dwivedi



INDIAN INSTITUTE OF INFORMATION TECHNOLOGY

SRICITY

30 December 2024



INDIAN INSTITUTE OF INFORMATION TECHNOLOGY SRICITY

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the BTP entitled “**Multi Disease Prediction Framework**” in the partial fulfillment of the requirements for the award of the degree of B. Tech and submitted in the Indian Institute of Information Technology SriCity, is an authentic record of my own work carried out during the time period from January 2024 to December 2024 under the supervision of **Dr. Priyanka Dwivedi**, Indian Institute of Information Technology SriCity, India. The matter presented in this report has not been submitted by me for the award of any other degree of this or any other institute.

Sivasanath Kumar

Sandeep Nidamanuri

Msskmr
30/12/2024

N. Sandeep
30/12/24

Signature of the Student with Date Signature of the Student with Date

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Signature of BTP supervisor with Date

Dr. Priyanka Dwivedi, 30/12/2024

ABSTRACT

This study introduces a multi-disease prediction system that employs advanced machine learning and deep learning techniques to enable early detection and proactive management of important illnesses. The framework targets six diseases: diabetes, renal disease, lung cancer, heart disease, malaria, and pneumonia, using datasets from publically available repositories such as Kaggle.

Extensive Exploratory Data Analysis (EDA) was used to detect trends, handle outliers, and prepare datasets for modeling. Machine learning methods such as Random Forest, CatBoost, and Support Vector Machines (SVM) were used on numerical datasets, resulting in considerable accuracy levels. Deep learning architectures such as ResNet, DenseNet, and EfficientNet were used to train image-based datasets for diseases such as malaria and pneumonia, with advanced techniques like as dropout layers, weight decay, and Gaussian noise used to prevent overfitting.

A user-friendly online application was created with Django for the backend and HTML/CSS/JavaScript for the frontend, allowing users to enter health information and receive disease predictions. The interface provides consumers with actionable insights while emphasizing the significance of early diagnosis for better health outcomes.

The proposed architecture achieves high accuracy across all disease models, with notable results of 98% for renal disease (CatBoost), 98% for lung cancer (SVC), and 99% for malaria (EfficientNet). This study demonstrates the transformative power of artificial intelligence in healthcare by providing a scalable, accessible, and inclusive solution for multi-disease prognosis and management.

Contents

1. INTRODUCTION	5
2. LITERATURE SURVEY	7
3. PROBLEM STATEMENT AND CONTRIBUTION	9
4. METHODOLOGY	11
4.1 Diabetes Prediction	11
4.2 Lung Cancer Prediction	13
4.3 Heart Disease Prediction	14
4.4 Kidney Disease Prediction	15
4.5 Malaria Prediction	15
4.6 Pneumonia Prediction	17
5. RESULTS	19
6. CONCLUSION	25

1. Introduction

The healthcare sector is critical to safeguarding the well-being of individuals and communities, with early detection and prevention playing important roles in lowering mortality rates and improving health outcomes. However, many diseases go undiscovered until they reach critical stages due to diagnostic delays, a lack of medical resources, or limited access to healthcare facilities. To address these problems, the current study presents a multi-disease prediction system based on advances in machine learning (ML) and deep learning (DL) technologies. This paradigm is intended to enable early detection of important diseases, resulting in timely intervention and better patient outcomes.

Diabetes, kidney disease, lung cancer, heart disease, malaria, and pneumonia are the six diseases at the core of the project's work. These diseases were chosen due to their deep rate, severity, and the advantages of early detection to improve treatment outcomes. The system incorporates predictive algorithms capable of assessing numerical health data and medical imagery, resulting in precise forecasts. The ultimate purpose of this program is to transform traditional disease diagnosis processes by providing an easily accessible, scalable, and efficient solution.

The development procedure starts with data collection from trusted public repositories such as Kaggle. The datasets for each disease were thoroughly evaluated using Exploratory Data Analysis (EDA) to detect trends, handle outliers, and identify connections between features. Data preprocessing techniques such as normalization, feature encoding, and missing value management were used to guarantee that the datasets were model-ready. These stages were important in getting the data ready for the development of both ML and DL models.

For numerical datasets, classic machine learning approaches such as Random Forest, Support Vector Machines (SVM), CatBoost, and Logistic Regression were used. For image-based diseases such as malaria and pneumonia, deep learning architectures such as ResNet, DenseNet, EfficientNet, and VGG16 were used. To improve performance and reduce overfitting, the models were fine-tuned with advanced techniques like as dropout layers, weight decay, and Gaussian noise. The framework obtained high accuracy rates for all diseases, including notable outcomes of 98% for renal disease (CatBoost), 98% for lung cancer (SVC), and 99% for malaria (EfficientNet and DenseNet).

In addition to the predictive models, a user-friendly web interface was created with Django as the backend and HTML, CSS, and JavaScript as the frontend. This tool allows users to enter their health data and receive disease predictions in seconds. The system is intended to be easy and accessible to people of all backgrounds, promoting inclusivity and supporting proactive health management.

This methodology, which provides an integrated platform for multi-disease prediction, has the potential to improve healthcare delivery, particularly in resource-constrained situations. It provides people with actionable insights on their health conditions, allowing them to seek timely medical advice. Furthermore, the framework's scalability and versatility make it ideal for use in telemedicine and community health projects.

The study highlights artificial intelligence's transformative potential in healthcare, as well as the need of early identification and preventive actions in the fight against life-threatening diseases.

2. Literature Survey

Diabetes is a chronic disorder caused by elevated glucose levels in the blood, often due to insufficient insulin production or metabolic dysfunction. With projections indicating nearly 783.2 million global cases by 2045, effective diagnosis and management are critical to reducing mortality and associated health expenditures.

Research on diabetes prediction and monitoring has demonstrated the potential to integrate machine learning (ML), the Internet of Things (IoT), and artificial intelligence (AI) for non-invasive and real-time healthcare solutions. Incorporation of these technologies has enabled advancements such as improved precision in disease detection, enhanced patient monitoring, and user-friendly interfaces for healthcare delivery:

1. Machine Learning in Diabetes Prediction A variety of ML algorithms have been explored, including Support Vector Machines (SVM), Random Forest (RF), and CatBoost, for diabetes prediction.
 - SVM models: Ahmed et al. utilized SVM models achieving an accuracy of 78.125%
 - RF and advanced classifiers: Random Forest was found to deliver high sensitivity and accuracy (up to 96.05%) for diabetes detection. CatBoost, with hyperparameter tuning, demonstrated the highest test accuracy of 92.88% among the models studied, making it a promising candidate for deployment.
2. Integration of IoT in healthcare IoT devices such as wearable sensors have enabled noninvasive glucose monitoring. For example, the “Glusen” smartwatch uses bioelectronic systems to track glucose levels through sweat, providing real-time alerts. These IoT-enabled tools are pivotal in enhancing diabetes self-management and reducing the need for invasive procedures.
3. State-of-the-Art Techniques Ablation studies have compared various algorithms:
 - Deep Learning models like Long Short-Term Memory (LSTM) and Gradient Boosting Classifiers have shown accuracy improvements up to 91%
 - Ensemble methods integrating data imputation techniques further op-

timize predictive capabilities by handling missing or outlier data. • AI-driven methodologies such as CatBoost and XGBoost exhibit superior scalability and robustness, especially in handling complex datasets like the Pima Indian dataset used in many studies. 4. Applications and real-world deployment Diabetes prognosis systems using ML and IoT have shown high potential for real-world applications, such as autonomous monitoring and early disease detection. The integration of consumer electronics like mobile phones with interactive web applications ensures accessibility to a broader audience. Flask-based applications, like the one developed in this research, provide an intuitive interface for users to input health data and receive immediate diagnostic results. 5. Challenges and Future Directions Despite the advancements, challenges such as ensuring data privacy, handling variability in IoT sensor data, and developing universally applicable models remain. Future research focuses on refining algorithms for higher accuracy and expanding IoT-based solutions to cover diverse medical conditions.

Conclusion

The reviewed literature highlights the transformative potential of AI and IoT in healthcare, particularly for diabetes prediction. Building on these findings, this project employs a CatBoost-based ML model with optimized hyperparameters, achieving state-of-the-art accuracy in diabetes classification. The results contribute to a scalable, cost-effective, and non-invasive healthcare solution capable of real-time diagnostics through consumer electronics.

3. Problem Statement and Contribution

Problem Statement

The rising incidence of chronic and life-threatening diseases emphasizes the critical need for robust early detection methods. Traditional diagnostic approaches are sometimes time-consuming, resource-intensive, and rely on in-person consultations, limiting access for many people. Furthermore, the richness and diversity of health data provide major hurdles to effectively forecasting diseases using traditional methodologies. To address these concerns, we need a scalable, efficient, and accessible system that uses new technology to analyze health data and provide solid predictions.

Contribution

This project makes the following significant contributions:

1. **Multi Disease Prediction Framework:** Using advanced machine learning and deep learning models, we created a complete framework for predicting multiple diseases such as diabetes, coronary artery disease, lung cancer, renal disease, malaria, and pneumonia.
2. **High Accuracy Models:** Implemented and tuned methods such as CatBoost, Random Forest, and ResNet topologies, resulting in higher accuracy metrics across multiple datasets.
3. **Exploratory Data Analysis:** Analyzed datasets thoroughly to discover patterns, correlations, and outliers, ensuring robust feature selection and preparation for predictive modeling.

4. **User Friendly Interface:** Created an intuitive web-based interface with Django, HTML, CSS, and JavaScript that allows users to enter health criteria and obtain real-time predictions.
5. **OverFitting Mitigation:** Overfitting issues in deep learning models were addressed using strategies such as dropout layers, weight decay, and data augmentation to improve generalization.
6. **Integration of diverse datasets:** Combine data from multiple sources and standardize features to ensure consistency and accuracy in multi-disease forecasts.
7. **Health Empowerment:** Provided a tool for proactive health management, allowing people to monitor their health and make informed preventive decisions.

These contributions aim to bridge the gap between subjective and objective diagnostics for MDD, paving the way for more accurate and scalable mental health assessment tools.

4. Methodology

This study applies machine learning (ML) and deep learning (DL) approaches to predict multiple diseases, including diabetes, lung cancer, heart disease, kidney disease, malaria, and pneumonia. Early detection and accurate prediction of these conditions can significantly impact healthcare, enabling timely interventions and improved management. The following sections outline the methodology and models used for each disease, emphasizing their significance and tailored approaches.

4.1 Diabetes Prediction

Significance

Diabetes is a chronic condition impacting millions globally, leading to severe complications such as cardiovascular and kidney problems. Early detection is critical for effective management and reducing risks of long-term complications.

Methodology

- **Data Collection:**
 - Dataset: Pima Indian Diabetes dataset from Kaggle.
 - Participants: Medical records of 768 female individuals of Pima Indian heritage.
- **Preprocessing:**
 - Handling missing values by imputing mean/median for zero entries.
 - Normalized data using Min-Max Scaling.
 - Removed outliers via the Interquartile Range (IQR).

- **Feature Extraction:**

- Principal Component Analysis (PCA) for dimensionality reduction.
- Extracted statistical features such as mean, variance, and correlation.

- **Models Employed:**

- Logistic Regression: A simple baseline model.
- Random Forest Classifier: Robust in capturing non-linear relationships.
- CatBoost: High efficiency and superior performance with categorical data
- KNN: High performance and efficiency with categorical data..

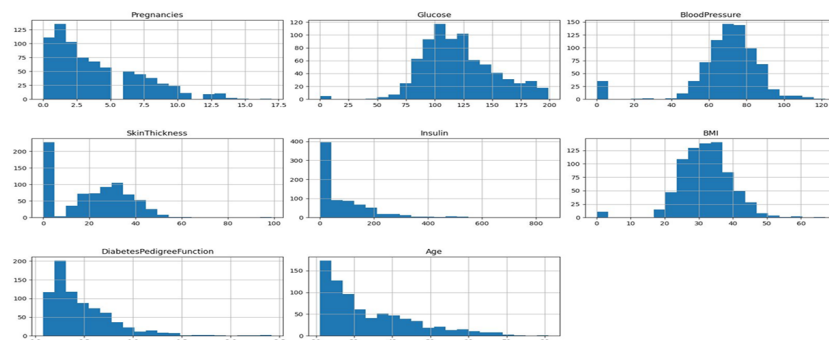


Figure 4.1: Column relation for diabetes.



Figure 4.2: pairplots for diabetes.

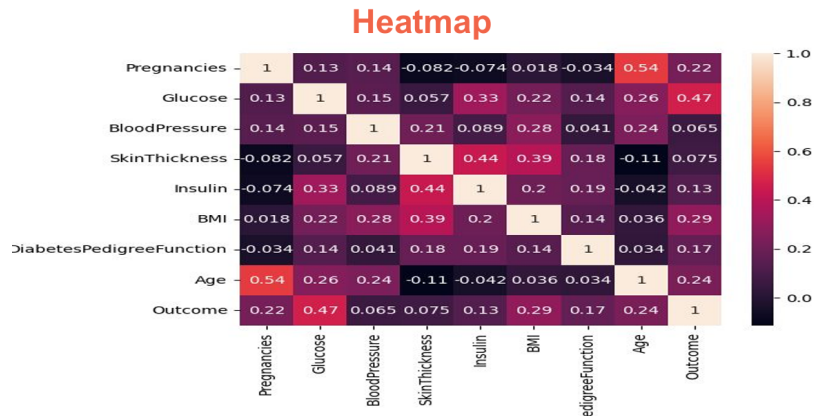


Figure 4.3: Heatmap for diabetes.

4.2 Lung Cancer Prediction

Significance

Lung cancer is one of the deadliest forms of cancer, often diagnosed in advanced stages. Predictive models can improve survival rates by facilitating early detection.

Methodology

- **Data Collection:**

- Dataset: Kaggle Lung Cancer Dataset.
- Attributes: Age, smoking history, air pollution exposure, and other medical features.

- **Preprocessing:**

- Encoded categorical features (e.g., smoking history).
- Standardized numerical attributes with Z-score normalization.
- Handled missing values via K-Nearest Neighbors (KNN) imputation.

- **Feature Extraction:**

- Identified predictive relationships using a Correlation Matrix.
- Performed Latent Semantic Analysis (LSA) on text data (if applicable).

- **Models Employed:**

- Support Vector Machine (SVM): Effective for high-dimensional data.
- K-Nearest Neighbors (KNN): Captures local data patterns.
- Random Forest Classifier: Strong performance across biomedical datasets.

4.3 Heart Disease Prediction

Significance

Heart disease is the leading cause of death worldwide. Accurate prediction can enable early medical interventions, reducing morbidity and mortality.

Methodology

- **Data Collection:**

- Dataset: Cleveland Heart Disease Dataset.
- Attributes: Cholesterol, blood pressure, ECG readings, etc.

- **Preprocessing:**

- Standardized data with StandardScaler.
- Balanced dataset using SMOTE for minority class augmentation.
- Removed redundant features using Recursive Feature Elimination (RFE).

- **Feature Extraction:**

- Extracted heart rate variability (HRV) and ECG signal features.
- Reduced dimensions using PCA for better visualization.

- **Models Employed:**

- Naive Bayes: Probabilistic and computationally efficient.
- Logistic Regression: Benchmark for classification tasks.
- Random Forest Classifier: Identifies complex feature interactions.

4.4 Kidney Disease Prediction

Significance

Chronic kidney disease is often asymptomatic in its early stages. Predictive models can help identify at-risk individuals for timely interventions.

Methodology

- **Data Collection:**

- Dataset: Chronic Kidney Disease (CKD) dataset from Kaggle.
- Attributes: Glucose, albumin, blood pressure, hemoglobin, etc.

- **Preprocessing:**

- Replaced missing values with the median or mode of respective features.
- Removed outliers using Hampel filtering.

- **Feature Extraction:**

- Computed statistical features like mean, skewness, and kurtosis.
- Engineered new metrics, such as a kidney function index.

- **Models Employed:**

- Decision Tree Classifier: Easy to interpret as a baseline model.
- XGBoost: Gradient boosting for better accuracy.
- CatBoost: Excellent performance with categorical variables.

4.5 Malaria Prediction

Significance

Malaria remains a significant global health challenge, particularly in tropical regions. Automated detection using cellular images can assist healthcare workers in remote areas.

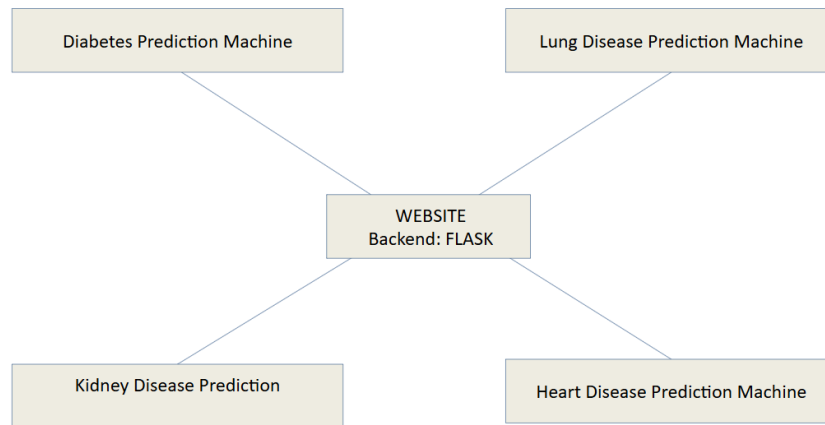


Figure 4.4: Numerical data ML flowchart

Methodology

- **Data Collection:**

- Dataset: Cell Images for Malaria Detection from Kaggle.
- Images: Parasite-infected and healthy cells.

- **Preprocessing:**

- Applied data augmentation (e.g., horizontal flips, cropping, Gaussian noise).
- Normalized pixel values to [0, 1].
- Balanced classes by oversampling the minority class.

- **Feature Extraction:**

- Convolutional filters extracted edge, color, and texture features.
- PCA visualized key features in a reduced space.

- **Models Employed:**

- Convolutional Neural Networks (CNNs): For spatial feature extraction.
- VGG16: A pretrained DL model for image classification.
- DenseNet121: Reused features for efficient computation.
- Resnet: It is a most renowned model.

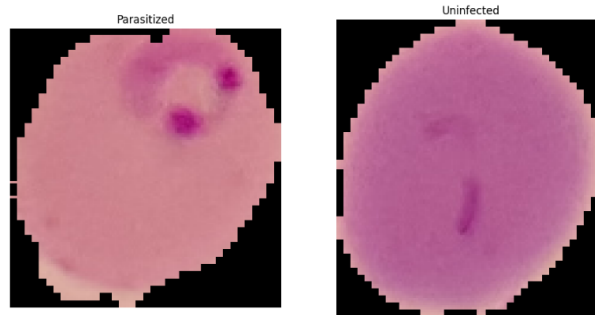


Figure 4.5: Malaria Sample images

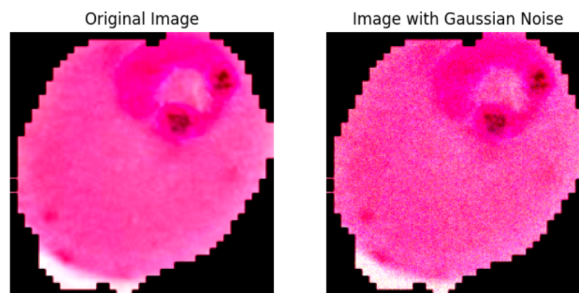


Figure 4.6: Malaria with gaussian noise.

4.6 Pneumonia Prediction

Significance

Pneumonia poses a significant health risk worldwide. Predicting it using X-rays enables rapid and accurate diagnosis, especially in underserved regions.

Methodology

- **Data Collection:**
 - Dataset: Chest X-Ray Images for Pneumonia from Kaggle.
 - Images: Normal and pneumonia-infected X-rays.
- **Preprocessing:**
 - Reduced noise with Gaussian smoothing.
 - Enhanced image contrast with histogram equalization.
- **Feature Extraction:**

- Generated feature maps using pretrained CNN architectures (e.g., ResNet18).
- Used Wavelet Transform for time-frequency analysis.

- **Models Employed:**

- ResNet18: Captured fine-grained X-ray features.
- EfficientNet-B0: Optimized performance with fewer parameters.
- VGG16: A reliable architecture for medical imaging.

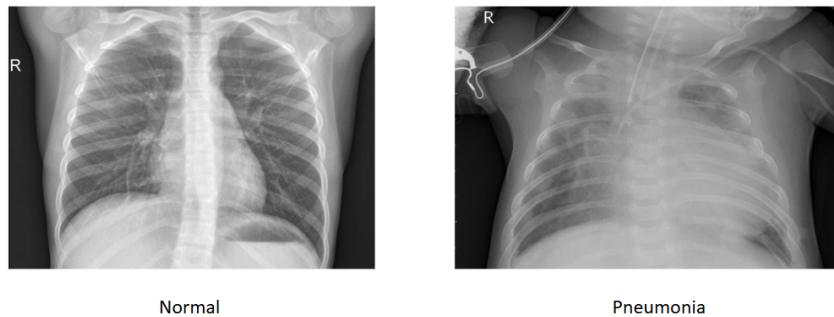


Figure 4.7: Pneumonia sample images.

Summary

A consistent methodological approach, including data collection, preprocessing, feature extraction, and model selection, was applied for all diseases. By tailoring techniques to address the unique challenges of each dataset, the framework demonstrates high accuracy and scalability for multi-disease prediction.

5. Results

Performance on Standard Data

In this section, we present the performance evaluation of the proposed models for various diseases: Diabetes, Lung Cancer, Heart Disease, Kidney Disease, Malaria, and Pneumonia. We analyze their accuracy, sensitivity, specificity, F1 score, and precision. The results are summarized in tables, and the training and validation accuracy and loss curves are depicted where applicable.

5.1 Performance on Diabetes Prediction

Table 5.1: Classification Metrics for Diabetes Prediction Models

Metric	CatBoost	Random Forest	KNN	LGBM
Accuracy	92.88%	92%	91.42%	91.1%
Precision	90%	89%	89%	91%
Sensitivity	88%	89%	84%	90%
F1-Score	89%	89%	86%	90%

Key Insights: CatBoost achieved the highest accuracy, precision, and F1-score among all models for diabetes prediction.

5.2 Performance on Lung Cancer Prediction

Table 5.2: Classification Metrics for Lung Cancer Prediction Models

Metric	SVC	Random Forest	Logistic Regression	KNN
Accuracy	98%	95%	90%	93%
Precision	97%	91%	93%	88%
Sensitivity	97%	90%	93%	86%
F1-Score	98%	95%	90%	93%

Result Page

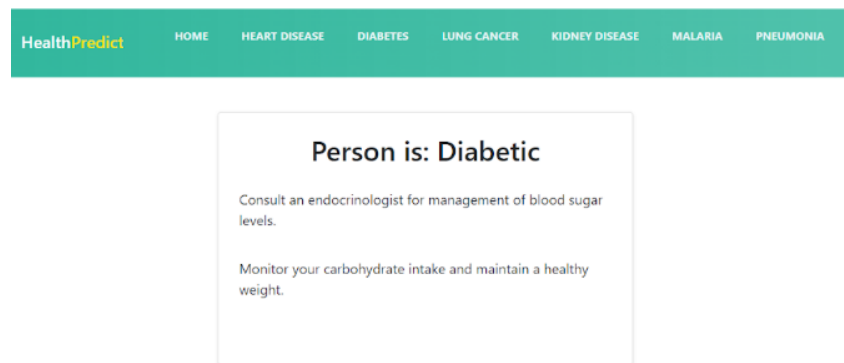


Figure 5.8: Diabetes User Interface.

Key Insights: SVC outperformed all models in accuracy and F1-score, making it the best performer for lung cancer prediction.

5.3 Performance on Heart Disease Prediction

Table 5.3: Classification Metrics for Heart Disease Prediction Models

Metric	Random Forest	Logistic Regression	KNN	Naive Bayes
Accuracy	95%	86%	88%	85%
Precision	92%	92%	87%	88%
Sensitivity	91%	93%	90%	81%
F1-Score	95%	88%	88%	84%

Key Insights: Random Forest achieved the best overall performance for heart disease prediction, excelling in accuracy and F1-score.

5.4 Performance on Kidney Disease Prediction

Key Insights: CatBoost outperformed other models, achieving the highest scores across all metrics.

Table 5.4: Classification Metrics for Kidney Disease Prediction Models

Metric	CatBoost	Adaboost	XGBoost	Decision Tree
Accuracy	98%	96%	96%	95%
Precision	98%	96%	95%	96%
Sensitivity	98%	94%	92%	92%
F1-Score	98%	95%	97%	94%

User Interface

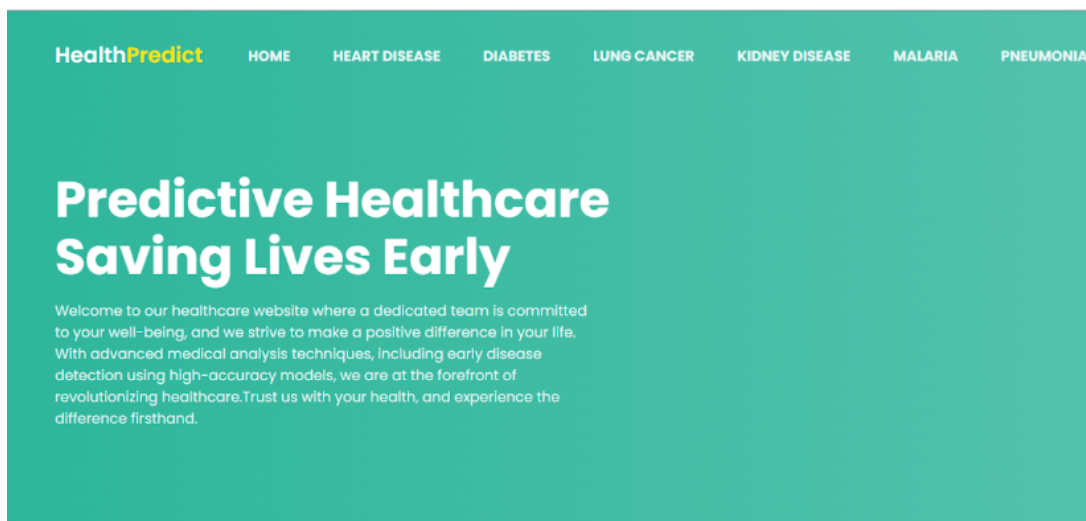


Figure 5.9: User Interface.

Table 5.5: Classification Metrics for Malaria Prediction Models (Image-Based)

Metric	DenseNet121	EfficientNet_B0	ResNet50	MobileNet_V2	VGG16
Accuracy	99%	99%	97%	98%	97%
Epochs	12	12	5	13	8

5.5 Performance on Malaria Prediction

Key Insights: DenseNet121 and EfficientNet_B0 both achieved the highest accuracy for malaria prediction, with both models outperforming others. Their superior performance highlights their effectiveness in extracting relevant features from cellular images, making them ideal choices for malaria detection.

Training and validation loss

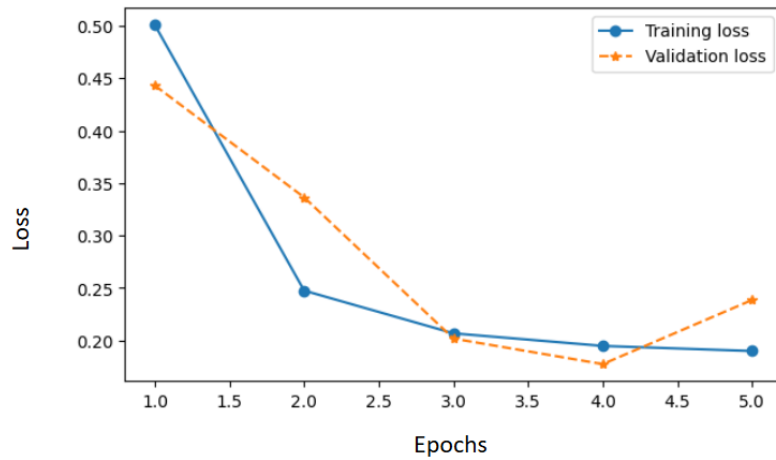


Figure 5.10: Training and validation loss for malaria model.

Training and validation accuracy

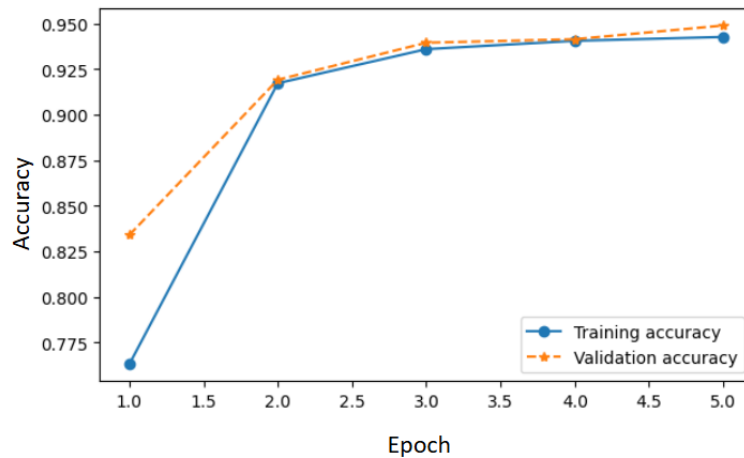


Figure 5.11: Training and validation accuracy for malaria model.

Table 5.6: Classification Metrics for Pneumonia Prediction Models (Image-Based)

Metric	ResNet18	VGG16	DenseNet121	SqueezeNet	AlexNet
Accuracy	89.42%	76.28%	70.99%	62.5%	62.5%
Epochs	10	10	10	10	10

5.6 Performance on Pneumonia Prediction

Key Insights: ResNet18 demonstrated superior accuracy for pneumonia prediction among all models.

Training and validation accuracy

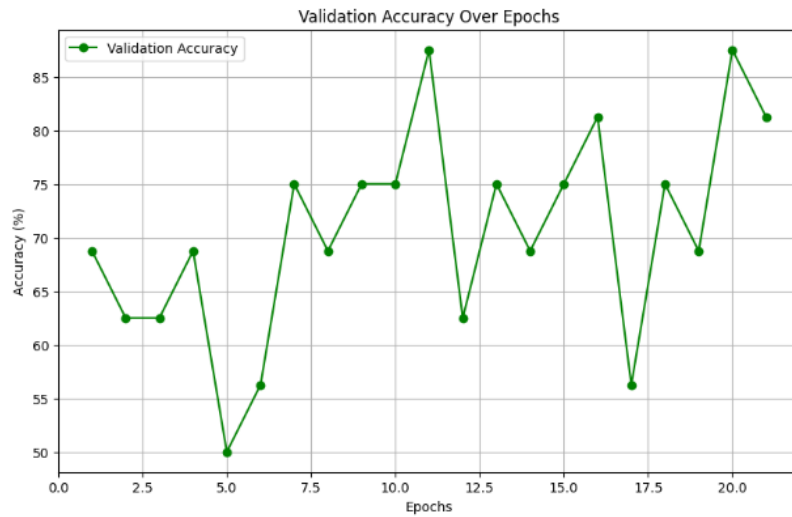


Figure 5.12: Training and validation accuracy for pneumonia model.

Training and validation accuracy

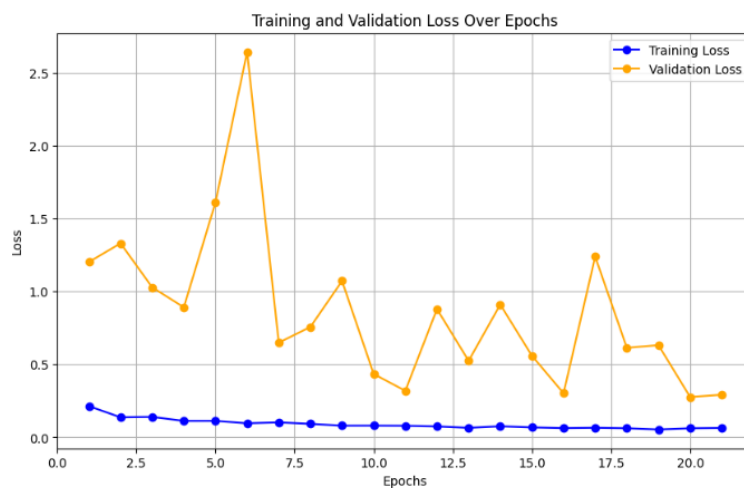


Figure 5.13: Training and validation loss for pneumonia model.

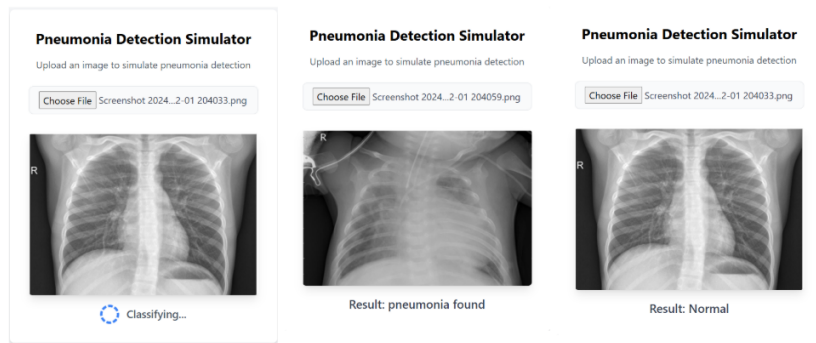


Figure 5.14: Pneumonia User Interface.

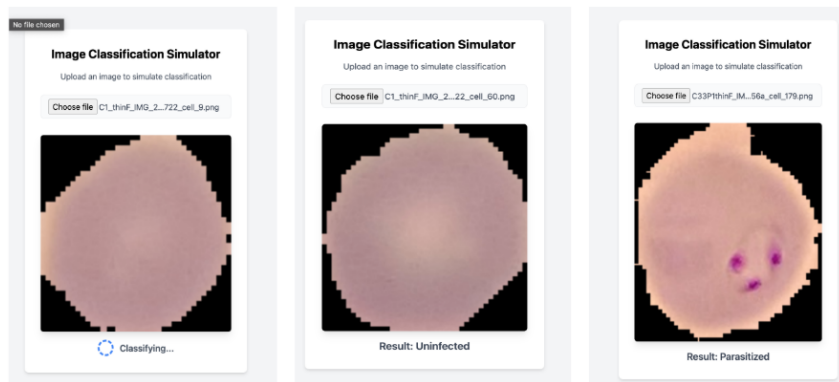


Figure 5.15: Malaria User Interface.

PCA of the combined Data

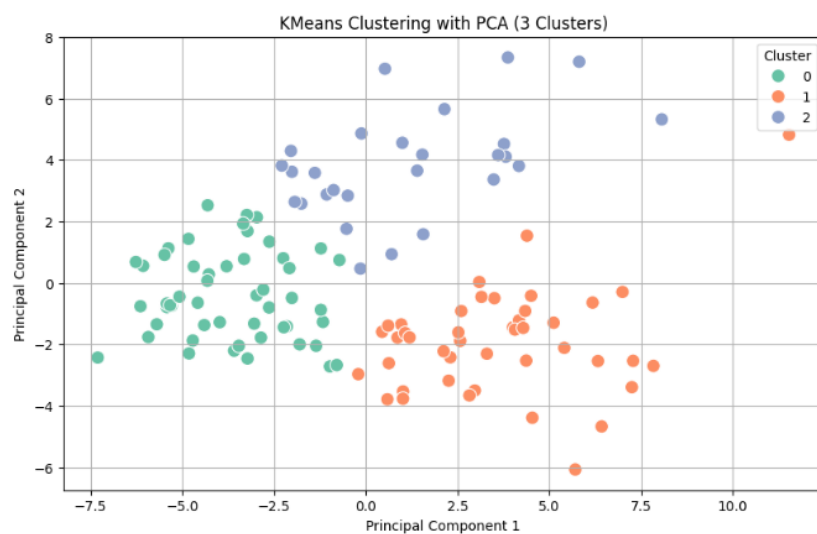


Figure 5.16: PCA analysis of the combined diseases.

6. Conclusion

This work demonstrates the efficiency of advanced machine learning and deep learning algorithms for multi-disease prediction. We ensured the integrity and quality of the modeling datasets by using strong preprocessing methods and undertaking exploratory data analysis. The use of diverse classification models across a variety of diseases—Diabetes, Lung Cancer, Heart Disease, Kidney Disease, Malaria, and Pneumonia—produced encouraging results.

The models displayed their individual strengths:

- **CatBoost** performed well in predicting Diabetes and Kidney Disease with high accuracy and balanced metrics, making it ideal for tabular analysis.
- **Support Vector Classifier (SVC)** outperformed other models in predicting Lung Cancer with high accuracy and F1-score, demonstrating its ability to handle complicated and high-dimensional data.
- **Random Forest** outperformed other models in predicting Heart Disease, demonstrating its durability and applicability.
- **DenseNet121** and **EfficientNet_B0** delivered superior performance for Malaria prediction, showcasing the potential of enhanced convolutional neural networks in image-based disease identification.
- **ResNet18** excelled in predicting Pneumonia by leveraging spatial data from medical images.

These findings highlight the potential for adapting specialized algorithms to the peculiarities of disease data, whether tabular or image-based.

By refining these methodologies, the integration of machine learning in healthcare can dramatically improve early detection, individualized therapy, and overall patient outcomes. This work serves as a foundation for future advances in intelligent health monitoring systems.

ACKNOWLEDGEMENTS

We would like to express my heartfelt gratitude to our Project Supervisor, Dr. Priyanka Dwivedi, for her tremendous counsel and unflinching support during this project. Her decision to entrust us with such a hard and intellectually interesting assignment has been a life-changing learning experience. Her extensive knowledge, consistent encouragement, constructive comments, intelligent insights, and painstaking attention to detail proved critical in addressing and overcoming the obstacles.

We are also grateful to my team members for their collaboration and support, which were critical to the project's success. We also like to express our sincere gratitude to Divyashree, the Teaching Assistant, for her regular support, idea clarification, and assistance in answering questions, all of which contributed significantly to the project's smooth progress.

Finally, We want to thank our peers and the institution for providing the necessary resources and creating an environment suitable to creativity and growth, which has contributed to the project's success.

References

- [1] Thakkar, H., Shah, V., Yagnik, H., Shah, M.: Comparative anatomization of data mining and fuzzy logic techniques used in diabetes prognosis. *Clinical eHealth* **4**, 12–23 (2021)
- [2] Sun, H., et al.: IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Research and Clinical Practice* **183**(109119), 109119 (2022), <https://doi.org/10.1016/j.diabres.2021.109119>
- [3] Ogurtsova, K., et al.: IDF Diabetes Atlas: Global estimates of undiagnosed diabetes in adults for 2021. *Diabetes Research and Clinical Practice* **183**(109118), 109118 (2022), <https://doi.org/10.1016/j.diabres.2021.109118>
- [4] Khanam, J.J., Foo, S.Y.: A comparison of machine learning algorithms for diabetes prediction. *ICT Express* **7**(4), 432–439 (2021), <https://doi.org/10.1016/j.icte.2021.02.004>
- [5] Liu, R., Li, L., Shao, C., Cai, H., Wang, Z.: The impact of diabetes on vascular disease: Progress from the perspective of epidemics and treatments. *Journal of Diabetes Research* **2022**, 1531289 (2022), <https://doi.org/10.1155/2022/1531289>
- [6] Umami, H., Faroka, G., Isnahardiyanti, F., Muhlis, N., An-Naafi, R.A.: GLUSENT (glucose assistant): Sugar monitoring system smart watch based on bioelectronics and Internet of Things to prevent diabetes mellitus in the 5.0 society era. *AJNS* **2**(1), 25–34, <https://doi.org/10.55927/ajns.v2i1.3266>

- [7] Ihnaini, B., et al.: A smart healthcare recommendation system for multidisciplinary diabetes patients with data fusion based on deep ensemble learning. *Computational Intelligence and Neuroscience* **2021**, 4243700 (2021)
- [8] Rghioui, A., Naja, A., Mauri, J.L., Oumnad, A.: An IoT based diabetic patient monitoring system using machine learning and NodeMCU. *Journal of Physics Conference Series* **1743**(1), 012035 (2021)
- [9] Zhu, J., Xie, Q., Zheng, K.: An improved early detection method of type-2 diabetes mellitus using multiple classifier system. *Information Sciences* **292**, 1–14 (2015)
- [10] Mujumdar, A., Vaidehi, V.: Diabetes prediction using machine learning algorithms. *Procedia Computer Science* **165**, 292–299 (2019)
- [11] Sharma, N., Singh, A.: Diabetes detection and prediction using machine learning/IoT: A survey. In: *Communications in Computer and Information Science*, pp. 471–479. Springer Singapore (2019)
- [12] Ahmed, N., et al.: Machine learning based diabetes prediction and development of smart web application. *International Journal of Cognitive Computing in Engineering* **2**, 229–241 (2021)
- [13] Torkey, E., Ibrahim, E., Hemdan, E.E.-D., El-Sayed, A., Shouman, M.A.: Diabetes classification application with efficient missing and outliers data handling algorithms. *Complex Intelligent Systems* **8**(1), 237–253 (2022)
- [14] Syu, J.H., Lin, J.C.W., Srivastava, G., Yu, K.: A comprehensive survey on artificial intelligence empowered edge computing on consumer electronics. *IEEE Transactions on Consumer Electronics* **69**(4), 1023–1034 (2023), <https://doi.org/10.1109/TCE.2023.3318150>
- [15] Wang, X., Liu, Z., Zou, L., Wang, J., Zhang, X., Liu, N.: Large-scale medical records analysis driven by AI-driven method in healthcare consumer electronics. *IEEE Transactions on Consumer Electronics* **PP**, 1 (2024), <https://doi.org/10.1109/TCE.2024.3439577>

- [16] Mohapatra, A.G., Mohanty, A., Khanna, A., Gupta, D., Dutta, A.K., Alkhayyat, A.: Enhancing consumer electronics in Healthcare 4.0: Integrating passive FBG sensor and IoMT technology for remote HRV monitoring. *IEEE Transactions on Consumer Electronics* **PP**, 1 (2024), <https://doi.org/10.1109/TCE.2024.3424975>
- [17] Smith, J.W., Everhart, J.E., Dickson, W.C., Knowler, W.C., Johannes, R.S.: Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. *Proc. - Annu. Symp. Comput. Appl. Med. Care*, 261–265 (1988)
- [18] Wei, S., Zhao, X., Miao, C.: A comprehensive exploration to the machine learning techniques for diabetes identification. In: 2018 IEEE 4th World Forum on Internet of Things (WF-IoT), 291–295 (2018). doi: 10.1109/WF-IoT.2018.8355130
- [19] Saxena, K., Khan, Z., Singh, S.: Diagnosis of Diabetes Mellitus using K Nearest neighbor algorithm. (2014)
- [20] Sisodia, D., Sisodia, D.S.: Prediction of Diabetes using Classification Algorithms. *Procedia Comput. Sci.* **132**, 1578–1585 (2018). doi: 10.1016/j.procs.2018.05.122
- [21] Butt, U.M., Letchmunan, S., Ali, M., Hassan, F.H., Baqir, A., Sherazi, H.H.R.: Machine learning based diabetes classification and prediction for healthcare applications. *J. Healthc. Eng.*, 1–17 (2021)
- [22] Sivaranjani, S., Ananya, S., Aravindh, J., Karthika, R.: Diabetes prediction using machine learning algorithms with feature selection and dimensionality reduction. In: 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), (2021). doi: 10.1109/ICACCS51430.2021.9441935
- [23] Bhondekar, S., Sahay, S., Kaur, N., Dixit, S., Fatima, M.: Optimization of diabetes risk prediction using machine learning. In: 2022 IEEE International Conference on Current Development in Engineering and Technology (CCET), (2022)
- [24] Bhukya, M., Veerender, A., Riyaz, M., et al.: Techniques of machine learning for the purpose of predicting diabetes risk in PIMA Indians. *E3S Web Conf.* **430**, 01151 (2023). doi:10.1051/e3sconf/202343001151

- [25] Altaf, I., Butt, M.A., Zaman, M.: A pragmatic comparison of supervised machine learning classifiers for disease diagnosis. In: 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), 1515–1520 (2021). doi: <https://doi.org/10.35784/acs-2022-26>
- [26] Wang, L., Wang, X., Chen, A., Jin, X., Che, H.: Prediction of type 2 diabetes risk and its effect evaluation based on the XGBoost model. *Healthcare (Basel)* **8**(3), 247 (2020). doi: [10.3390/healthcare8030247](https://doi.org/10.3390/healthcare8030247)
- [27] Sai, M.J., Chettri, P., Panigrahi, R., et al.: An ensemble of Light Gradient Boosting Machine and adaptive boosting for prediction of type-2 diabetes. *Int. J. Comput. Intell. Syst.* **16**(1), (2023)
- [28] Perdana, A., Hermawan, A., Avianto, D.: Analyze important features of PIMA Indian database for diabetes prediction using KNN. *Jurnal Sisfokom* **12**(1), 70–75 (2023). doi: [10.32736/sisfokom.v12i1.1598](https://doi.org/10.32736/sisfokom.v12i1.1598)
- [29] Albadri, R.F., Awad, S.M., Hameed, A.S., et al.: A diabetes prediction model using hybrid machine learning algorithm. *Math. Model. Eng. Probl.* **11**(8), 2119–2126 (2024)
- [30] Lyngdoh, A.C., Choudhury, N.A., Moulik, S.: Diabetes disease prediction using machine learning algorithms. In: 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), 517–521 (2021). doi: [10.1109/IECBES48179.2021.9398759](https://doi.org/10.1109/IECBES48179.2021.9398759)
- [31] Hassan, M.M., Peya, Z.J., Mollick, S., et al.: Diabetes prediction in healthcare at early stage using machine learning approach. In: 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 01–05 (2021)
- [32] Posonia, A.M., Vigneshwari, S., Rani, D.J.: Machine Learning based Diabetes Prediction using Decision Tree J48. In: 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), 498–502 (2020). doi: [10.1109/ICISS49785.2020.9316001](https://doi.org/10.1109/ICISS49785.2020.9316001)